

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОПРЕДЕЛЕНИЯ МАКСИМАЛЬНО ВОЗМОЖНОГО ЗНАЧЕНИЯ НАВЕДЕННОГО НАПРЯЖЕНИЯ НА ОТКЛЮЧЕННОЙ ЛИНИИ ЭЛЕКТРОПЕРЕДАЧИ

1 Постановка задачи

При отключении линии электропередачи (ЛЭП) на ней токами влияющих ЛЭП (соседние работающие ЛЭП) неизбежно наводится напряжение некоторого значения. При превышении наведенного напряжения предельно допустимого значения 25 В возникает опасность поражения персонала, проводящего работы на отключенной ЛЭП, электрическим током.

Отсюда следует постановка задачи: определить в заданной точке отключенной ЛЭП максимально возможное значение модуля наведенного напряжения при изменении фазных токов влияющих ЛЭП в пределах диапазонов их возможных значений.

2 Теоретические основы наведенного напряжения

Из решения уравнений Максвелла для электромагнитного поля, следует, что наведенное напряжение на отключенной ЛЭП определяется следующим выражением (подчеркивание означает, что величина относится к множеству \mathbb{C} комплексных чисел)

$$\underline{u} = \sum_{i=1}^N \left(\underline{I}_{An_i} \underline{Z}_{An_i} + \underline{I}_{Bn_i} \underline{Z}_{Bn_i} + \underline{I}_{Cn_i} \underline{Z}_{Cn_i} + \underline{I}_{Ak_i} \underline{Z}_{Ak_i} + \underline{I}_{Bk_i} \underline{Z}_{Bk_i} + \underline{I}_{Ck_i} \underline{Z}_{Ck_i} \right), \quad (1)$$

где N – число влияющих ЛЭП;

$\underline{I}_{An_i}, \underline{I}_{Bn_i}, \underline{I}_{Cn_i}, \underline{I}_{Ak_i}, \underline{I}_{Bk_i}, \underline{I}_{Ck_i}$ – фазные токи в начале и конце i -ой влияющей ЛЭП;

$\underline{Z}_{An_i}, \underline{Z}_{Bn_i}, \underline{Z}_{Cn_i}, \underline{Z}_{Ak_i}, \underline{Z}_{Bk_i}, \underline{Z}_{Ck_i}$ – взаимные сопротивления между отключенной ЛЭП и фазными проводами в начале и конце i -ой влияющей ЛЭП.

Выражение (1) является уравнением линейной регрессии, коэффициенты которой – взаимные сопротивления $\underline{Z}_{An_i}, \underline{Z}_{Bn_i}, \underline{Z}_{Cn_i}, \underline{Z}_{Ak_i}, \underline{Z}_{Bk_i}, \underline{Z}_{Ck_i}$ априори неизвестны. При этом, в общем случае, каждая влияющая ЛЭП имеет шесть предикторов (признаков) – фазные токи $\underline{I}_{An_i}, \underline{I}_{Bn_i}, \underline{I}_{Cn_i}, \underline{I}_{Ak_i}, \underline{I}_{Bk_i}, \underline{I}_{Ck_i}$ в ее начале и конце.

С помощью метода симметричных составляющих выражение (1) можно преобразовать к следующему виду

$$\underline{u} = \sum_{i=1}^N \left(\underline{I}_{1n_i} \underline{Z}_{1n_i} + \underline{I}_{2n_i} \underline{Z}_{2n_i} + 3\underline{I}_{0n_i} \underline{Z}_{0n_i} + \underline{I}_{1k_i} \underline{Z}_{1k_i} + \underline{I}_{2k_i} \underline{Z}_{2k_i} + 3\underline{I}_{0k_i} \underline{Z}_{0k_i} \right), \quad (2)$$

$$\underline{I}_{1n_i} = \frac{\underline{I}_{An_i} + \underline{I}_{Bn_i} \underline{a} + \underline{I}_{Cn_i} \underline{a}^2}{3}, \quad \underline{I}_{1k_i} = \frac{\underline{I}_{Ak_i} + \underline{I}_{Bk_i} \underline{a} + \underline{I}_{Ck_i} \underline{a}^2}{3}, \quad (3)$$

$$\underline{I}_{2n_i} = \frac{\underline{I}_{An_i} + \underline{I}_{Bn_i} \underline{a}^2 + \underline{I}_{Cn_i} \underline{a}}{3}, \quad \underline{I}_{2k_i} = \frac{\underline{I}_{Ak_i} + \underline{I}_{Bk_i} \underline{a}^2 + \underline{I}_{Ck_i} \underline{a}}{3}, \quad (4)$$

$$\underline{I}_{0n_i} = \frac{\underline{I}_{An_i} + \underline{I}_{Bn_i} + \underline{I}_{Cn_i}}{3}, \quad \underline{I}_{0k_i} = \frac{\underline{I}_{Ak_i} + \underline{I}_{Bk_i} + \underline{I}_{Ck_i}}{3}, \quad (5)$$

$$\underline{Z}_{1n_i} = \underline{Z}_{An_i} + \underline{Z}_{Bn_i} \underline{a}^2 + \underline{Z}_{Cn_i} \underline{a}, \quad \underline{Z}_{1k_i} = \underline{Z}_{Ak_i} + \underline{Z}_{Bk_i} \underline{a}^2 + \underline{Z}_{Ck_i} \underline{a}, \quad (6)$$

$$\underline{Z}_{2n_i} = \underline{Z}_{An_i} + \underline{Z}_{Bn_i} \underline{a} + \underline{Z}_{Cn_i} \underline{a}^2, \quad \underline{Z}_{2k_i} = \underline{Z}_{Ak_i} + \underline{Z}_{Bk_i} \underline{a} + \underline{Z}_{Ck_i} \underline{a}^2, \quad (7)$$

$$\underline{Z}_{0n_i} = \frac{\underline{Z}_{An_i} + \underline{Z}_{Bn_i} + \underline{Z}_{Cn_i}}{3}, \quad \underline{Z}_{0k_i} = \frac{\underline{Z}_{Ak_i} + \underline{Z}_{Bk_i} + \underline{Z}_{Ck_i}}{3}, \quad (8)$$

где $\underline{a} = e^{j2\pi/3}$ – оператор поворота на комплексной плоскости;

$\underline{I}_{1n_i}, \underline{I}_{2n_i}, \underline{I}_{0n_i}, \underline{I}_{1k_i}, \underline{I}_{2k_i}, \underline{I}_{0k_i}$ – токи прямой, обратной и нулевой последовательностей в начале и конце i -ой влияющей ЛЭП;

$\underline{Z}_{1n_i}, \underline{Z}_{2n_i}, \underline{Z}_{0n_i}, \underline{Z}_{1k_i}, \underline{Z}_{2k_i}, \underline{Z}_{0k_i}$ – взаимные сопротивления прямой, обратной и нулевой последовательностей между отключенной ЛЭП и фазными проводами в начале и конце i -ой влияющей ЛЭП.

Определить максимально возможное значение модуля наведенного напряжения u_{\max} можно по следующему выражению для его точной верхней границы [1]

$$u_{\max} = \sum_{i=1}^N \left(I_{1n_{\max_i}} Z_{1n_i} + I_{2n_{\max_i}} Z_{2n_i} + 3I_{0n_{\max_i}} Z_{0n_i} + \right. \\ \left. + I_{1k_{\max_i}} Z_{1k_i} + I_{2k_{\max_i}} Z_{2k_i} + 3I_{0k_{\max_i}} Z_{0k_i} \right), \quad (9)$$

где $I_{1n_{\max_i}}, I_{2n_{\max_i}}, 3I_{0n_{\max_i}}, I_{1k_{\max_i}}, I_{2k_{\max_i}}, 3I_{0k_{\max_i}}$ – максимально возможные значения модулей токов симметричных составляющих в начале и конце i -ой влияющей ЛЭП.

Допустим, что значения $I_{1n_max_i}$, $I_{2n_max_i}$, $3I_{0n_max_i}$, $I_{1k_max_i}$, $I_{2k_max_i}$, $3I_{0k_max_i}$ известны для всех влияющих ЛЭП. Тогда из выражения (9) следует, что задача определения максимально возможного значения наведенного напряжения u_{max} сводится к задаче определения неизвестных коэффициентов линейной регрессии – взаимных сопротивлений Z_{1ni} , Z_{2ni} , Z_{0ni} , Z_{1ki} , Z_{2ki} , Z_{0ki} прямой, обратной и нулевой последовательностей.

В частном случае, когда длина влияющей ЛЭП составляет менее 100 км, фазные токи в ее начале можно принять равными соответствующим фазным токам в ее конце. Тогда каждая влияющая ЛЭП будет иметь не шесть, а только три предиктора – фазные токи \underline{I}_{Ai} , \underline{I}_{Bi} , \underline{I}_{Ci} или их симметричные составляющие I_{1i} , I_{2i} , I_{0i} . В этом случае выражения (1), (2), (9) принимают следующий вид

$$\underline{u} = \sum_{i=1}^N (\underline{I}_{Ai} \underline{Z}_{Ai} + \underline{I}_{Bi} \underline{Z}_{Bi} + \underline{I}_{Ci} \underline{Z}_{Ci}) = \sum_{i=1}^N (\underline{I}_{1i} \underline{Z}_{1i} + \underline{I}_{2i} \underline{Z}_{2i} + 3\underline{I}_{0i} \underline{Z}_{0i}), \quad (10)$$

$$u_{max} = \sum_{i=1}^N (I_{1_max_i} Z_{1i} + I_{2_max_i} Z_{2i} + 3I_{0_max_i} Z_{0i}), \quad (11)$$

где \underline{Z}_{Ai} , \underline{Z}_{Bi} , \underline{Z}_{Ci} – взаимные сопротивления между отключенной ЛЭП и фазными проводами i -ой влияющей ЛЭП;

\underline{Z}_{1i} , \underline{Z}_{2i} , \underline{Z}_{0i} – взаимные сопротивления прямой, обратной и нулевой последовательностей между отключенной ЛЭП и фазными проводами i -ой влияющей ЛЭП;

$I_{1_max_i}$, $I_{2_max_i}$, $3I_{0_max_i}$ – максимально возможные значения модулей токов симметричных составляющих i -ой влияющей ЛЭП.

3 Метод определения значения u_{max}

Для определения достоверного максимально возможного значения наведенного напряжения u_{max} предлагается использовать следующий метод, в основе которого лежат три операции.

Первая операция – проведение на отключенной ЛЭП (до проведения на ней работ) в течении некоторого интервала времени серии из M синхронных измерений комплексных значений наведенного напряжения и фазных токов влияющих ЛЭП.

Вторая операция – определение неизвестных коэффициентов линейной регрессии Z_{1ni} , Z_{2ni} , Z_{0ni} , Z_{1ki} , Z_{2ki} , Z_{0ki} посредством машинного обучения на результатах натурных измерений наведенного напряжения.

Третья операция – вычисление значения u_{max} по выражению (9) или (11).

Следует отметить, что необходимое число измерений для построения достоверной математической модели наведенного напряжения посредством машинного обучения может исчисляться десятками и сотнями тысяч.

Результатом натурных измерений является датасет (набор данных) из M объектов (результатов измерений), содержащий комплексные числа. Целевой переменной является наведенное напряжение \underline{u} , предикторами являются фазные токи $\underline{I}_{An_i}, \underline{I}_{Bn_i}, \underline{I}_{Cn_i}, \underline{I}_{Ak_i}, \underline{I}_{Bk_i}, \underline{I}_{Ck_i}$ влияющих ЛЭП.

В данной работе способы идентификации ЛЭП в качестве влияющих не рассматриваются. Принимается, что все влияющие ЛЭП известны. Следует отметить, что реальное число влияющих ЛЭП может исчисляться десятками и, следовательно, общее число предикторов может составить несколько сотен.

4 Особенности датасета с результатами натурных измерений

Полученный в результате натурных измерений датасет может иметь следующие особенности.

Первой особенностью датасета является возможное скачкообразное изменение целевой переменной – наведенного напряжения при изменении токовых нагрузок влияющих ЛЭП или направления передачи энергии по ним. Следствием этого является наличие в датасете кластеров данных, соответствующих периодам с разными нагрузками ЛЭП и направлениями передачи энергии. При этом некоторые кластеры данных могут соответствовать случаю передачи энергии по влияющим ЛЭП в противоположных направлениях, что приводит к существенному снижению значений наведенного напряжения на отключенной ЛЭП. Причиной этого является взаимная компенсация составляющих наведенного напряжения от разных влияющих ЛЭП.

Обучение на данных с малыми значениями наведенного напряжения может привести к занижению коэффициентов регрессии у обученных моделей и, следовательно, к занижению прогноза максимально возможного значения наведенного напряжения u_{\max} (так, например, нулевые значения целевой переменной приводят к тривиальному решению, то есть нулевым коэффициентам регрессии). Поэтому обучение на наборе данных, состоящем из нескольких кластеров, может привести к построению недостоверной модели регрессии наведенного напряжения. Отсюда следует вывод, что для определения достоверного значения u_{\max} необходимо проводить обучение обособленных моделей для каждого кластера данных с последующим анализом и отбором построенных моделей.

Второй особенностью датасета является возможная мультиколлинеарность предикторов – фазных токов, обусловленная одинаковыми трендами токовых нагрузок влияющих ЛЭП. Следствием этой мультиколлинеарности является неустойчивость обученной модели регрессии, устранить которую можно либо отбором предикторов, либо использованием модели с регуляризацией. Однако, необходимо понимать, что токи влияющих ЛЭП, как правило, не имеют функциональной зависимости и их мультиколлинеарность обусловлена временной корреляцией их значений, присущей некоторым периодам токовых нагрузок ЛЭП. Возможно, что для набора данных, полученного в другой период проведения натурных измерений, корреляция предикторов будет пренебрежимо мала. Отсюда следует вывод, что для определения достоверного значения u_{\max} не стоит исключать какие-либо предикторы из модели. При этом для обеспечения устойчивости модели следует использовать модели с регуляризацией.

5 Специфические особенности задачи определения значения u_{\max}

Поставленная задача определения максимально возможного значения наведенного напряжения по своей сути является задачей экстраполяции регрессии комплексных чисел. Данная задача имеет специфические особенности, которые накладывают ограничения на использование стандартных методов машинного обучения.

Первая особенность данной задачи заключается в принципиальной невозможности использования для экстраполяции регрессии широко используемых ансамблевых алгоритмов на основе деревьев решений и опорных векторов, так как за пределами обучающего набора данных модели на этих алгоритмах являются константными. Для решения задачи экстраполяции следует использовать алгоритмы линейной регрессии или нейронные сети.

Вторая особенность данной задачи заключается в том, что стандартные алгоритмы машинного обучения не поддерживают работу с комплексными числами. Это обстоятельство не позволяет использовать популярные библиотеки машинного обучения для построения модели регрессии на данных, содержащих комплексные числа. Отсюда возникает постановка вспомогательной задачи: выполнить математически обоснованное преобразование исходного датасета с комплексными числами в датасет с вещественными числами, которое позволит для построения модели регрессии комплексных чисел использовать стандартные

алгоритмы машинного обучения, поддерживающие работу только с вещественными числами.

6 Решение вспомогательной задачи

Матричное уравнение для определения неизвестных коэффициентов регрессии имеет следующий вид

$$\underline{I} \underline{z} = \underline{u}, \quad (12)$$

где \underline{I} – матрица измеренных комплексных значений предикторов – фазных токов влияющих ЛЭП;

\underline{z} – неизвестный вектор комплексных коэффициентов регрессии – взаимных сопротивлений между отключенной и влияющими ЛЭП;

\underline{u} – вектор измеренных комплексных значений целевой переменной – наведенного напряжения на отключенной ЛЭП.

В общем случае значения предикторов и целевой переменной заданы в виде датасета размером $M \times (6N + 1)$. При этом матрица \underline{I} имеет размер $M \times 6N$, а вектор \underline{u} имеет размер $M \times 1$.

Представим матричное уравнение (12) в следующем виде

$$(I_x + jI_y)(z_x + jz_y) = u_x + ju_y, \quad (13)$$

где индексами x и y обозначены, соответственно, реальные и мнимые части комплексных величин, затем раскрываем левую часть (13)

$$I_x z_x - I_y z_y + j(I_y z_x + I_x z_y) = u_x + ju_y. \quad (14)$$

Далее разделяя действительную и мнимую части (14), получаем систему из двух вещественных матричных уравнений

$$\begin{cases} I_x z_x - I_y z_y = u_x \\ I_y z_x + I_x z_y = u_y \end{cases}, \quad (15)$$

которую можно представить в компактной матричной форме

$$I z = u, \quad (16)$$

где I – расширенная вещественная матрица размером $2M \times 12N$, элементами которой являются вещественные матрицы I_x и I_y измеренных значений предикторов

$$I = \begin{pmatrix} I_x & -I_y \\ I_y & I_x \end{pmatrix}; \quad (17)$$

z – расширенный вещественный вектор размером $12N \times 1$, элементами которого являются векторы z_x и z_y неизвестных коэффициентов регрессии

$$z = \begin{pmatrix} z_x \\ z_y \end{pmatrix}; \quad (18)$$

u – расширенный вещественный вектор размером $2M \times 1$, элементами которого являются векторы u_x и u_y измеренных значений целевой переменной

$$u = \begin{pmatrix} u_x \\ u_y \end{pmatrix}. \quad (19)$$

Таким образом математическая модель линейной регрессии комплексных чисел, коэффициенты которой определяются матричным уравнением (12), сводится к модели линейной регрессии вещественных чисел, коэффициенты которой определяются уравнением (16). Для машинного обучения этой модели регрессии вещественных чисел необходимо вместо исходного датасета размером $M \times (6N + 1)$ с комплексными числами использовать преобразованный датасет размером $2M \times (12N + 1)$ с вещественными числами, полученный путем объединения вещественной матрицы I и вещественного вектора u .

Прогноз u_{pred} целевой переменной для некоторого заданного вектора размером $6N \times 1$ комплексных значений предикторов $\underline{I}_{pred} = I_{pred_x} + jI_{pred_y}$ описывается операторным выражением

$$u_{pred} = I_{pred}^T z, \quad (20)$$

где I_{pred} – матрица размером $12N \times 2$

$$I_{pred} = \begin{pmatrix} I_{pred_x} & I_{pred_y} \\ -I_{pred_y} & I_{pred_x} \end{pmatrix}. \quad (21)$$

Для получения прогноза I_{pred} целевой переменной необходимо в обученную модель передать матрицу I_{pred}^T (верхний индекс Т означает транспонирование матрицы). Результат u_{pred} прогноза целевой переменной будет представлен в виде вектора размером 2×1 , в котором первый его элемент является реальной частью комплексного числа \underline{u}_{pred} , а второй элемент – мнимой частью \underline{u}_{pred} .

7 Алгоритмы, используемые для обучения моделей регрессии

7.1 Алгоритмы из библиотеки Scikit-learn

Обучение моделей в данной работе выполнено на языке программирования Python 3. Для построения моделей использовались следующие алгоритмы линейной регрессии из библиотеки Scikit-learn – LinearRegression, использующий метод наименьших квадратов, а также Ridge, SGDRegressor, Lasso и BayesianRidge, использующие различные виды регуляризации.

Оптимальные значения гиперпараметров моделей регрессии определялись двумя способами – поиском по сетке и применением алгоритма Optuna.

Нейронные сети для обучения моделей в данной работе не использовались.

7.2 Алгоритм регуляризации Тихонова с использованием априорной информации

Алгоритм регуляризации Тихонова [2] с использованием априорной информации о погрешностях измерения предикторов и целевой переменной позволяет получить устойчивое решение для плохообусловленных систем линейных алгебраических уравнений (СЛАУ) с данными, заданными с некоторой погрешностью. В данной задаче наведенного напряжения мультиколлинеарность предикторов как раз и приводит к плохой обусловленности СЛАУ (16). При этом измерения значений целевой переменной и предикторов всегда проводятся с некоторыми погрешностями, значения которых предполагаем известными.

Отличие алгоритма регуляризации Тихонова от алгоритма Ridge из Scikit-learn состоит в том, что коэффициент регуляризации α определяется по условию оптимизации нестандартной метрики – обобщенной невязки, учитывающей погрешности измерения предикторов и целевой переменной [2].

Решение задачи регрессии (16) для алгоритма регуляризации Тихонова имеет следующий вид

$$z = (I^T I + \alpha E)^{-1} I^T u. \quad (22)$$

Уравнение обобщенной невязки для определения коэффициента регуляризации α имеет следующий вид [2]

$$\|Iz(\alpha) - u\|^2 - (h\|z(\alpha)\| + d)^2 = 0 \quad (23)$$

$$h = \delta_A \|I\| \quad (24)$$

$$d = \delta_b \|u\| \quad (25)$$

где $\|\cdot\|$ – обозначение нормы матрицы или вектора;

δ_A – относительная погрешность измерения элементов матрицы I ;

δ_b – относительная погрешность измерения элементов вектора u .

Таким образом в регуляризации Тихонова используется априорная информация о погрешностях измерения данных, определяемая выражениями (24), (25). В качестве дополнительной априорной информации в данной работе использовались известные из теории наведенного напряжения соотношения для модулей взаимных сопротивлений прямой, обратной и нулевой последовательностей

$$z_{1i} \approx z_{2i}; \quad z_{0i} \gg z_{1i}; \quad z_{0i} \gg z_{2i}. \quad (26)$$

8 Пайплайн машинного обучения для определения значения u_{\max}

На рисунке 1 приведен пайплайн (последовательность операций) определения максимально возможного значения наведенного напряжения посредством машинного обучения на результатах натурных измерений наведенного на отключенной ЛЭП напряжения и фазных токов влияющих ЛЭП.



Рисунок 1 – Пайплайн машинного обучения для определения максимально возможного значения наведенного напряжения

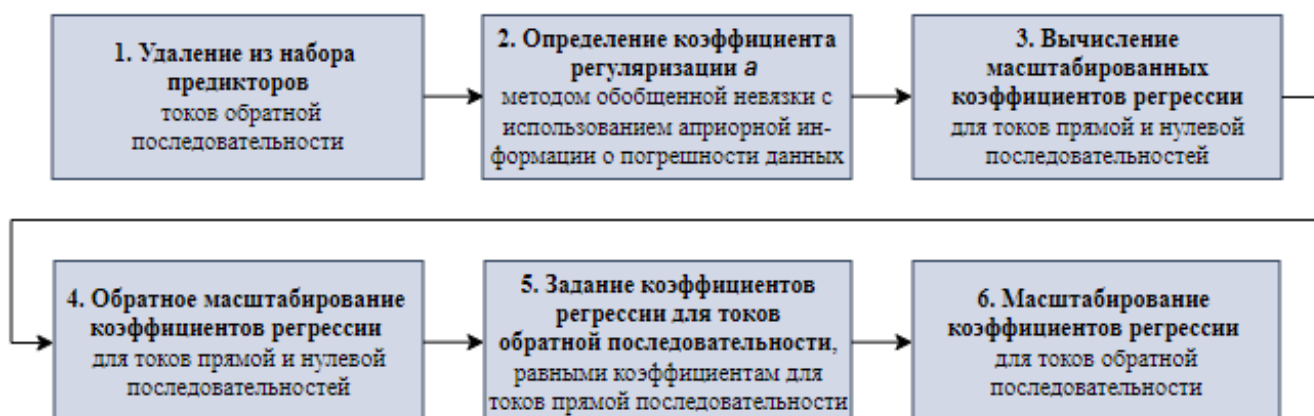


Рисунок 2 – Пайплайн машинного обучения модели регуляризации Тихонова

Операция машинного обучения модели регуляризации Тихонова с использованием априорной информации (24) – (26) более подробно приведена в виде отдельного пайплайна на рисунке 2.

9 Пример определения значения u_{\max}

Проведение серийных измерений значений наведенного на отключенной ЛЭП напряжения и фазных токов влияющих ЛЭП пока еще не практикуется. Поэтому для примера решения поставленной задачи использовались синтетические наборы данных, содержащие по 1000 объектов, имитирующих результаты измерений.

Для этого сначала был сгенерирован базовый датасет, который содержит комплексные значения фазных токов пяти влияющих ЛЭП и комплексные значения наведенного напряжения, вычисленные по выражению (1) при значениях взаимных сопротивлений, приведенных в таблице 1.

Начальные значения амплитуд фазных токов влияющих ЛЭП составляют: 200 А; 125 А; 220 А; 300 А; 170 А, соответственно. Далее происходит монотонное возрастание трендов модулей фазных токов каждой влияющей ЛЭП.

Базовый датасет содержит три кластера данных, соответствующих скачкообразным изменениям нагрузок влияющих ЛЭП и направлениям передачи энергии. Первое изменение нагрузок смоделировано скачкообразным изменением направления тока в ЛЭП №1 на 700 А, второе изменение смоделировано скачкообразным изменением направления тока в ЛЭП №3 на 800 А – см. рис. 3. Число объектов в кластерах составляет 350, 250 и 400, соответственно.

В результате базовый датасет представляет собой функциональную зависимость целевой переменной u от фазных токов в началах и концах влияющих ЛЭП.

Затем путем случайных отклонений предикторов в пределах 1% и целевой переменной u в пределах 5% от соответствующих значений из базового датасета генерировались датасеты, моделирующие результаты натурных измерений наведенного напряжения и фазных токов влияющих ЛЭП. Таким образом каждый сгенерированный датасет представляет собой одну и ту же серию измерений с случайными отклонениями измеренных значений наведенного напряжения и фазных токов в пределах погрешностей их измерений, равных $\delta_A = 0,01$ – для фазных токов и $\delta_B = 0,05$ – для наведенного напряжения.

Наибольшее для всех сгенерированных датасетов значение модуля целевой переменной – наведенного напряжения составило 6,7 В.

Таблица 1 – Взаимные сопротивления ЛЭП для генерации базового датасета

№ ЛЭП	Взаимные сопротивления ЛЭП, Ом					
	Z_{An_i}	Z_{Bn_i}	Z_{Cn_i}	Z_{Ak_i}	Z_{Bk_i}	Z_{Ck_i}
1	$0.040 + j0.0060$	$0.038 + j0.0055$	$0.036 + j0.0050$	$0.035 + j0.0048$	$0.033 + j0.0045$	$0.030 + j0.0044$
2	$0.020 + j0.0025$	$0.018 + j0.0022$	$0.016 + j0.0020$	$0.015 + j0.0019$	$0.014 + j0.0016$	$0.012 + j0.0014$
3	$0.018 + j0.0022$	$0.016 + j0.0020$	$0.015 + j0.0018$	$0.013 + j0.0016$	$0.012 + j0.0015$	$0.011 + j0.0014$
4	$0.016 + j0.0020$	$0.014 + j0.0018$	$0.013 + j0.0016$	$0.010 + j0.0015$	$0.009 + j0.0014$	$0.008 + j0.0012$
5	$0.015 + j0.0018$	$0.014 + j0.0016$	$0.013 + j0.0015$	$0.011 + j0.0012$	$0.010 + j0.0012$	$0.009 + j0.0010$

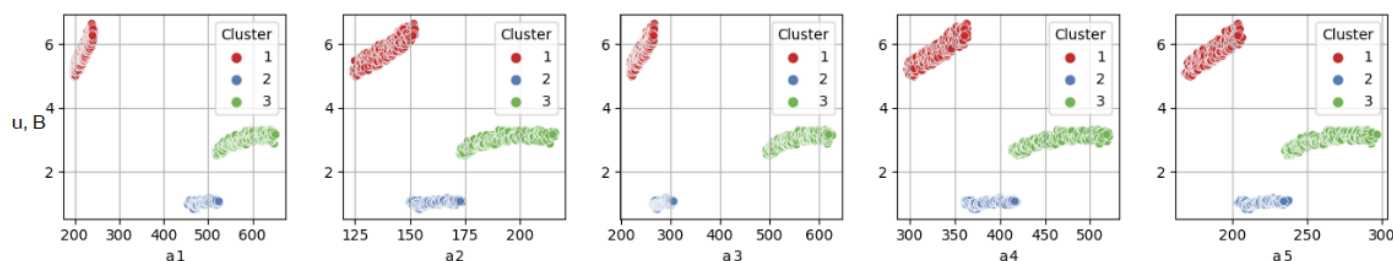


Рисунок 3 – Диаграммы рассеяния модуля наведенного напряжения от модулей токов в фазе А влияющих ЛЭП для dataset_1

Максимально возможное значение наведенного напряжения u_{\max} определялось для следующих максимально возможных значений токов влияющих ЛЭП (максимально возможные значения токов в началах влияющих ЛЭП приняты равными соответствующим токам в концах ЛЭП).

Токи прямой последовательности: 1500 А; 1200 А; 1600 А; 1600 А; 1200 А.

Токи обратной последовательности: 150 А; 120 А; 160 А; 160 А; 120 А.

Токи нулевой последовательности: 97,5 А; 78 А; 104 А; 104 А; 78 А.

Обучение моделей проводилось при разбиении данных на обучающую и тестовую выборки в соотношении 80/20.

Модели строились для каждого выявленного кластера данных в отдельности с использованием всех предикторов, которые масштабировались относительно наибольшего по модулю значения методом MaxAbsScaler. Программный код машинного обучения на языке Python 3 выложен в репозитории автора [3].

В таблицах 2 – 6 приведены результаты расчетов для пяти сгенерированных датасетов – значения коэффициентов детерминации R^2 и прогнозы значений u_{\max} различными алгоритмами линейных регрессий из библиотеки Scikit-learn. Окончания «GS» и «Opt» у названий алгоритмов в таблицах означают, что гиперпараметры моделей оптимизированы, соответственно, поиском по сетке и с помощью алгоритма Optuna.

Там же приведены результаты прогнозов для алгоритма регуляризации Тихонова (Regularized) и условно точный прогноз (Accurate), вычисленный по выражению (9) и коэффициентам регрессии из таблицы 1.

Таблица 2 – Метрики и результаты прогноза целевой переменной для dataset_1
(базовый датасет: random_seed_1 = 2022, random_seed_2 = 1148)

Кластер 1

	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.995	
R2 test	0.999	0.999	0.999	0.998	0.998	0.999	0.999	0.999	0.999	0.993	
U_max	111.0	66.2	65.7	111.5	103.8	43.9	40.8	65.9	65.7	83.4	91.6

Кластер 2

	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.993	0.993	0.992	0.988	0.988	0.992	0.991	0.992	0.992	0.991	
R2 test	0.987	0.987	0.986	0.984	0.983	0.986	0.987	0.986	0.986	0.986	
U_max	149.1	116.0	28.5	28.4	26.8	22.5	14.9	28.5	31.5	20.7	91.6

Кластер 3

	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.998	0.998	0.997	0.995	0.996	0.997	0.997	0.997	0.997	0.995	
R2 test	0.997	0.997	0.997	0.995	0.995	0.997	0.997	0.997	0.997	0.996	
U_max	195.3	160.7	48.9	63.5	60.5	41.4	44.5	48.8	48.9	30.3	91.6

Таблица 3 – Метрики и результаты прогноза целевой переменной для dataset_2
(базовый датасет: random_seed_1 = 2022, random_seed_2 = 1986)

Кластер 1

	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.995	
R2 test	0.999	0.999	0.999	0.998	0.998	0.999	0.999	0.999	0.999	0.995	
U_max	230.4	65.0	64.8	105.5	107.2	60.3	39.1	64.8	64.7	84.4	91.6

Кластер 2

	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.993	0.992	0.992	0.987	0.987	0.992	0.992	0.992	0.992	0.99	
R2 test	0.988	0.987	0.986	0.983	0.983	0.986	0.986	0.986	0.986	0.987	
U_max	118.5	67.9	27.4	30.7	29.9	22.4	25.2	28.2	27.7	20.5	91.6

Кластер 3

	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.998	0.998	0.997	0.995	0.995	0.997	0.997	0.997	0.997	0.995	
R2 test	0.997	0.997	0.997	0.996	0.996	0.997	0.997	0.997	0.997	0.996	
U_max	204.7	191.2	48.1	60.7	60.9	42.1	36.9	56.8	48.1	30.1	91.6

Таблица 4 – Метрики и результаты прогноза целевой переменной для dataset_3
(базовый датасет: random_seed_1 = 2022, random_seed_2 = 1987)

Кластер 1

	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.995	
R2 test	0.999	0.999	0.999	0.998	0.998	0.999	0.999	0.999	0.999	0.995	
U_max	299.8	207.6	75.2	102.0	103.3	52.2	39.4	73.0	73.0	81.1	91.6

Кластер 2

	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.991	0.991	0.99	0.986	0.986	0.99	0.99	0.99	0.99	0.989	
R2 test	0.987	0.987	0.988	0.987	0.987	0.987	0.987	0.987	0.988	0.986	
U_max	136.0	115.7	24.4	28.2	29.0	17.4	15.8	26.0	25.5	20.4	91.6

Кластер 3

	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.997	0.997	0.997	0.996	0.995	0.997	0.997	0.997	0.997	0.995	
R2 test	0.998	0.998	0.998	0.996	0.996	0.998	0.998	0.998	0.998	0.996	
U_max	145.7	109.0	50.5	63.9	62.1	41.1	40.3	50.5	50.5	31.3	91.6

Таблица 5 – Метрики и результаты прогноза целевой переменной для dataset_4
(базовый датасет: random_seed_1 = 2022, random_seed_2 = 2012)

Кластер 1											
	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.995	
R2 test	0.999	0.999	0.999	0.998	0.999	0.999	0.999	0.999	0.999	0.994	
u_max	180.8	71.7	70.1	108.0	113.4	52.6	50.4	70.2	70.1	85.7	91.6
Кластер 2											
	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.991	0.99	0.99	0.986	0.986	0.99	0.99	0.99	0.991	0.989	
R2 test	0.991	0.989	0.99	0.983	0.983	0.989	0.99	0.99	0.99	0.988	
u_max	126.7	21.8	25.1	27.2	27.4	13.3	23.0	23.8	25.8	18.0	91.6
Кластер 3											
	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.997	0.997	0.997	0.995	0.995	0.997	0.997	0.997	0.997	0.995	
R2 test	0.997	0.997	0.997	0.995	0.995	0.997	0.997	0.997	0.997	0.995	
u_max	179.5	178.8	49.7	58.6	58.5	39.8	35.8	56.5	49.6	30.6	91.6

Таблица 6 – Метрики и результаты прогноза целевой переменной для dataset_5
(базовый датасет: random_seed_1 = 2022, random_seed_2 = 1953)

Кластер 1											
	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.995	
R2 test	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.995	
u_max	204.4	125.2	70.5	82.8	85.8	42.4	41.2	70.0	70.6	83.4	91.6
Кластер 2											
	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.992	0.991	0.991	0.987	0.987	0.99	0.989	0.991	0.991	0.99	
R2 test	0.985	0.985	0.985	0.983	0.984	0.986	0.986	0.985	0.985	0.987	
u_max	105.9	24.8	24.0	29.1	28.5	15.5	10.1	25.9	25.3	18.2	91.6
Кластер 3											
	LinearRegr	Ridge_GS	Ridge_Opt	SGDRegr_GS	SGDRegr_Opt	Lasso_GS	Lasso_Opt	BayesRidge_GS	BayesRidge_Opt	Regularized	Accurate
R2 train	0.998	0.998	0.997	0.995	0.995	0.997	0.997	0.997	0.997	0.995	
R2 test	0.997	0.997	0.996	0.995	0.995	0.997	0.997	0.997	0.996	0.995	
u_max	193.5	181.0	50.9	63.9	61.1	45.1	41.7	58.3	50.9	31.7	91.6

10 Анализ результатов определения значения u_{\max}

Условно точный прогноз Accurate максимально возможного значения наведенного напряжения u_{\max} , вычисленный по коэффициентам регрессии из таблицы 1, составляет 91,6 В. Значение Accurate будем использовать в качестве эталонного значения для проверки точности прогнозирования различными моделями.

Наиболее близкими к условно точному прогнозу Accurate являются прогнозы моделей, обученных на данных кластера №1. Данные из кластера №1 соответствуют случаю, когда энергия по влияющим ЛЭП передается в одном направлении. Данные кластеров №2 и №3 соответствуют случаю, когда энергия по некоторым влияющим ЛЭП передается в разных направлениях, что приводит к

определению заниженных коэффициентов моделей регрессии и, как следствие, к занижению прогноза наведенного напряжения.

В таблице 7 приведены средние значения (mean), смещения (bias) и среднеквадратичные отклонения ($\sqrt{\text{variance}}$) прогнозов для рассмотренных датасетов, выполненные различными алгоритмами, обученными на данных кластеров №1.

Таблица 7 – Значения mean, bias и $\sqrt{\text{variance}}$ для прогнозов различных алгоритмов (датасеты dataset_1 - dataset_5; кластер №1)

	Linear Regress	Ridge GS	Ridge Optuna	SGD GS	SGD Optuna	Lasso GS	Lasso Optuna	Bayes GS	Bayes Optuna	Regular Tikhonov	Accurate
mean, B	205,3	107,1	69,3	102,0	102,7	50,3	42,9	68,8	68,8	83,6	91,6
bias, B	113,7	15,5	-22,3	10,4	11,1	-41,3	-48,7	-22,8	-22,8	-8,0	0
bias / Accurate	1,24	0,17	-0,24	0,11	0,12	-0,45	-0,53	-0,25	-0,25	-0,09	0
$\sqrt{\text{variance}}$, B	69,0	61,5	4,2	11,3	10,3	7,3	6,3	3,4	3,5	1,7	0
$\sqrt{\text{variance}}$ / mean	0,34	0,57	0,06	0,11	0,10	0,14	0,15	0,05	0,05	0,02	0

Несмотря на отличное значение коэффициента детерминации R^2 для всех алгоритмов регрессий из библиотеки Scikit-learn (см. таблицы 2 – 6), результаты их прогнозов не только имеют неприемлемую точность, но и являются неустойчивыми – сравнительно небольшие погрешности измерений предикторов и целевой переменной $\delta_A = 0,01$ и $\delta_B = 0,05$ приводят к недопустимым погрешностям прогноза целевой переменной (в терминологии машинного обучения это означает, что модели переобучаются). Для рассмотренных датасетов смещения прогнозов различных алгоритмов регрессий составляют (по модулю) от 11% до 124%, а среднеквадратичные отклонения – от 5% до 57%. Наихудшие прогнозы для алгоритмов Scikit-learn выдает не использующий регуляризацию алгоритм LinearRegression, наилучшие прогнозы – алгоритм SGDRegressor.

Наилучшие устойчивые прогнозы с приемлемой точностью выдает алгоритм регуляризации Тихонова, учитывающий априорную информацию о погрешностях измерения данных. Для рассмотренных датасетов смещение прогнозов алгоритма регуляризации Тихонова составляет (по модулю) 9%, а среднеквадратичное отклонение – 2%.

Выводы

1. Для определения достоверного максимально возможного значения наведенного напряжения на отключенной ЛЭП разработан метод, в основе которого лежит построение математической модели регрессии наведенного напряжения посредством машинного обучения на данных, полученных путем натурных измерений наведенного напряжения и фазных токов влияющих ЛЭП.

2. Полученный в результате натурных измерений набор данных может иметь кластеры, соответствующие периодам с разными нагрузками влияющих ЛЭП и направлениями передачи энергии. Для определения достоверного максимально возможного значения наведенного напряжения необходимо проводить обучение обособленных моделей для каждого кластера данных с последующим анализом и отбором построенных моделей.

3. Стандартные алгоритмы машинного обучения не поддерживают работу с комплексными числами, которыми описывается наведенное напряжение, что не позволяет использовать популярные библиотеки машинного обучения для построения модели регрессии на данных, содержащих комплексные числа.

4. Разработан метод преобразования исходного набора данных с комплексными числами в набор данных с вещественными числами, который позволяет для построения модели регрессии комплексных чисел использовать стандартные алгоритмы машинного обучения, поддерживающие работу только с вещественными числами.

5. Модели, использующие алгоритмы линейной регрессии из библиотеки Scikit-learn, обученные на наборах данных с мультиколлинеарными предикторами – фазными токами, могут привести к крайне неточным и неустойчивым прогнозам наведенного напряжения.

6. Для наборов данных с мультиколлинеарными предикторами для получения приемлемо точных и устойчивых прогнозов наведенного напряжения следует использовать модель регуляризации Тихонова, учитывающую априорную информацию о погрешностях измерения предикторов и целевой переменной.

Список литературы

1. Горшков А. В. Определение максимального значения наведенного напряжения в рассматриваемой точке отключенной воздушной линии электропередачи. – Электричество, 2017, № 11. – С. 12 - 21.

2. Тихонов А. Н., Гончарский А. В., Степанов В.В., Ягола А. Г. Регуляризирующие алгоритмы и априорная информация. – М.: Наука, 1983. – 200 с.

3. <https://github.com/GorshkovAndrey/Prediction-of-the-maximum-value-of-induced-voltage-on-a-disconnected-power-line>